MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

DEVELOPMENT AND COMPARISON OF M-ESTIMATORS FOR LOCATION
ON THE BASIS OF THE ASYMPTOTIC VARIANCE FUNCTIONAL

by

Dennis D. Boos and R. J. Serfling

FSU Statistics Report M380
ONR Technical Report No. 108

Technical rept.,   ONR   TR-108

July 1976
Department of Statistics
The Florida State University
Tallahassee, Florida 32306

33 p.

ABSTRACT

DEVELOPMENT AND COMPARISON OF M-ESTIMATORS FOR LOCATION
ON THE BASIS OF THE ASYMPTOTIC VARIANCE FUNCTIONAL

A new approach for comparison and modification of M-estimators is intro-
duced and implemented. The problem considered is that of robust estimation
of a location parameter. Specific attention is given to the ϵ-contaminated
normal model. The analytical method introduced is based upon the asymptotic
variance of the estimator, the asymptotic variance being considered as a
functional over the space of distribution functions. The behavior of this
functional is investigated with respect to special sub-families within a
neighborhood of the "central" distribution. With respect to the normal loca-
tion problem, three robust estimators from the 1972 Princeton Monte Carlo
study are examined: the "Huber" H15, the "Hampel" 25A, and the sine
function M-estimator AMT. Using the asymptotic variance functional analysis
as both an analytical and intuitive tool, three modified estimators are sug-
gested and developed. All six estimators are then compared at selected
distributions heavier tailed than the normal. Besides its analytical and
intuitive appeal, this functional approach offers a cost-saving alternative
to Monte Carlo methods.

1. INTRODUCTION

The notion of M-estimation was introduced by Huber [8] in order to develop
a robust estimator for the mean of a normal distribution. The approach has
since played a central role in robustness theory for location estimation and
has been directed toward other estimation problems as well. For a theoretical
overview see Huber [9] and for a broad Monte Carlo study see Andrews *et al.*
[1]. Some features of M-estimation apropos to the motives of the present
investigation are:

> functional characterizations of parameters are natural;
> computations via iterative methods are relatively easy;
> fine tuning of the estimators is straightforward; the
> *M-estimators* form an "admissible" class in a sense made
> precise by Hampel [7, p. 391].

The present paper introduces a new approach for comparison of M-estimators.
It is in character an analytical tool, based on asymptotic variance consid-
ered as a functional, but also the approach facilitates the use of intuition
and provides a cost-saving alternative to Monte Carlo methods. We utilize
the approach to illuminate some previous M-estimators and to develop new ones.
Attention is confined to the location problem with scale known.

Let us define what is meant by an M-estimator for a location parameter
in the case of scale known. Consider i.i.d. observations $X_1, \cdots, X_n$ from a
distribution $F_\theta(x) = F(x - \theta)$, for some distribution F. In this case θ is a
real-valued location parameter. Let ψ be any function such that θ is the
solution T of the equation

$$(1.1) \qquad \int \psi(x - T) dF_\theta(x) = 0 .$$

Then a natural estimator of θ is given by the solution $\hat{\theta}$ of the equation

(1.2) $$\int \psi(x - \hat{\theta})dF_n(x) = 0 ,$$

where F_n denotes the usual sample distribution function. We call this the *M-estimator* of θ corresponding to ψ and sometimes designate the estimator by referring to the function ψ rather than the solution $\hat{\theta}$. (For some choices of ψ, the equations (1.1) and (1.2) may have multiple solutions. We assume implicitly that in such a case there is a procedure for selecting a single one of the solutions as appropriate in the role of θ or $\hat{\theta}$.) Note that (1.1) implies the relation

(1.3) $$\int \psi(x)dF(x) = 0 \quad \left(= \int \psi(x - \theta)dF(x - \theta) \right)$$

between ψ and F. Note also that θ takes the value 0 at the distribution F.

If F has mean 0, in which case the location parameter θ may be represented as the mean of F_θ, then (1.1) is fulfilled for $\psi(x) = x$ and the corresponding M-estimator given by (1.2) is the sample mean. Note that in this case the equation (1.2) for $\hat{\theta}$ is equivalent to the equation for obtaining $\hat{\theta}$ by the method of least squares. If F is symmetric with unique median 0, then the M-estimator corresponding to $\psi(x) = -1, 0,$ or 1 according as $x < 0, = 0,$ or > 0, is the sample median and is equivalent to estimation by the method of least absolute values. If $F = \alpha G + (1 - \alpha)H$, where $0 \leq \alpha \leq 1$, G is absolutely continuous with density g, and H is discrete with mass function h, and if $f = \alpha g + (1 - \alpha)h$ is differentiable, then for $\psi = f'/f$ the solution of (1.2) is the maximum likelihood estimator of θ. Note that this choice of ψ depends upon the particular distribution F generating the location model.

Under suitable regularity conditions on ψ and F, $\sqrt{n}\,(\hat{\theta} - \theta)$ is asymptotically normal with mean 0 and variance (independent of θ since θ is a location parameter) given by

$$(1.4) \qquad A(F) = \frac{\int \psi^2(x)\,dF(x)}{[\int \psi'(x)\,dF(x)]^2} \ .$$

(See, for example, Rao [10, p.378] or Huber [8].) Clearly, A(F) may be viewed as a functional defined on a class of possible distribution functions F. This characterization of the asymptotic variance is helpful in two ways: it facilitates looking at A(F) for small departures from F, and it leads into the use of existing theory of functionals in the analysis of A(F).

The notion of considering the *estimator* $\hat{\theta}$ as a statistical functional $T(F_n)$, based on the representation $\theta = T(F_\theta)$ expressed by (1.1), is an established approach developed by von Mises [11], further developed by Filippova [5] and Hampel [6], [7], and exploited in Andrews *et al.* [1]. The novelty of the present investigation is that the "functional approach" is directed not toward the estimator but rather toward the asymptotic variance of the estimator. In a similar vein, we note that Bickel and Lehmann [2, p. 1058] have expressed interest in notions of continuity of A(F) in regard to the problem of superefficiency of estimators.

Basically, we want to find estimators of θ which are asymptotically efficient at Φ, the standard normal, but which also protect against long-tailed contamination. In Section 2 we develop the basic techniques for analyzing, by the functional approach, the asymptotic variance A(W) of an M-estimator as W ranges through an ϵ-neighborhood of the form

$$\{W: W = (1 - \epsilon)F_0 + \epsilon G \ , \ G \text{ symmetric}\} \ .$$

Here F_o is a fixed "central" distribution. In the sequel we take F_o to be Φ . The techniques of Section 2 are utilized in Section 3 to shed new light on several known estimators and to set up guidelines for building new estimators. In Section 4 we develop three new estimators, which are compared in Section 5 with the known estimators of Section 3. As brought out in Section 5, these analytical techniques for comparison of estimators not only have intuitive appeal but also provide a competitor to the Monte Carlo approach. General-izations and ramifications are considered in Section 6.

2. THE ASYMPTOTIC VARIANCE FUNCTIONAL

We consider M-estimation for a location parameter, as described in Section 1. Related to a particular function ψ is an asymptotic variance functional A(F), given by (1.4), defined on distribution functions F. This characterization of the asymptotic variance parameter of the M-estimator based on ψ allows us to draw upon the powerful functional approach introduced in the statistical context by von Mises [11].

DEFINITION 2.1. A functional T is *m times differentiable* at a point (distribution) F with respect to a convex set τ of distributions if for each $W \epsilon \tau$ and for each $t \epsilon [0, 1]$, the derivative

$$(2.1) \qquad \frac{d^m}{dt^m} T[F + t(W - F)]$$

exists.

Corresponding to this special differentiability, we have a Taylor expansion.

THEOREM. *Let $T(\cdot)$ be m times differentiable at F with respect to a set τ. Let $W \in \tau$. Then there exists $z \in (0, 1)$ such that*

(2.2)

$$T(W) - T(F) = \sum_{p=1}^{m-1} \frac{1}{p!} \frac{d^p}{dt^p} T[F + t(W - F)]\Big|_{t = 0} + \frac{1}{m!} \frac{d^m}{dt^m} T[F + t(W - F)]\Big|_{t = z}.$$

In order to apply this theorem, we compute the first two derivatives of the asymptotic variance functional $A(F)$. In carrying out these computations, we assume that F is symmetric about 0, that ψ is skew-symmetric $\big(\psi(-x) = - \psi(x)\big)$, and that certain differentiations may be taken under the integral sign. We obtain

$$\frac{d}{dt} A[F + t(W - F)] = \frac{\int \psi^2 d(W-F)}{[\int \psi' dF + t\int \psi' d(W-F)]^2}$$

(2.3)
$$- \frac{2[\int \psi' d(W - F)][\int \psi^2 dF + t\int \psi^2 d(W - F)]}{[\int \psi' dF + t\int \psi' d(W - F)]^3}$$

and

$$\frac{d^2}{dt^2} A[F + t(W - F)] = - \frac{4[\int \psi' d(W - F)][\int \psi^2 d(W - F)]}{[\int \psi' dF + t\int \psi' d(W - F)]^3}$$

(2.4)
$$+ \frac{6[\int \psi^2 dF + t\int \psi^2 d(W - F)][\int \psi' d(W - F)]^2}{[\int \psi' dF + t\int \psi' d(W - F)]^4}.$$

Setting $t = 0$ in (2.3), we obtain

$$(2.5) \quad \frac{d}{dt} A[F + t(W - F)]\Big|_{t = 0} = A(F)[1 - \frac{2\int \psi' dW}{\int \psi' dF}] + \frac{\int \psi^2 dW}{[\int \psi' dF]^2}.$$

Denote this derivative $\Delta_{F,W}$. Likewise (2.4) yields

$$\frac{d^2}{dt^2} A[F + t(W - F)]\Big|_{t = 0} = 2A(F)\{1 - 4\frac{\int \psi' dW}{\int \psi' dF} + 3[\frac{\int \psi' dW}{\int \psi' dF}]^2\}$$

$$\text{(2.6)} \qquad + 4\frac{\int \psi^2 dW}{[\int \psi' dF]^2} [1 - \frac{\int \psi' dW}{\int \psi' dF}] .$$

Denote this by $\Delta_{F,W}^{(2)}$. Clearly we could continue taking derivatives and obtain the following expression for the asymptotic variance of a distribution W:

$$\text{(2.7)} \qquad A(W) = A(F) + \Delta_{F,W} + \tfrac{1}{2} \Delta_{F,W}^{(2)} + 1/6 \Delta_{F,W}^{(3)} + \cdots .$$

Our interest lies in distributions W which are "close" to the base distribution F. Suppose that we take $W = (1 - \epsilon)F + \epsilon G$, where G can be viewed as a contaminating distribution. This particular characterization of a contamination situation has been used by Huber [8] and others and possesses desirable mathematical properties. In this case

$$\Delta_{F,W} = \frac{d}{dt} A[F + t((1 - \epsilon)F + \epsilon G - F))]\Big|_{t = 0}$$

$$= \frac{d}{dt} A[F + \epsilon t(G - F)]\Big|_{t = 0}$$

$$\text{(2.8)} \qquad = \epsilon \Delta_{F,G} .$$

Similarly, $\quad \Delta_{F,W}^{(2)} = \epsilon^2 \Delta_{F,G}^{(2)}$, etc. Thus

$$\text{(2.9)} \qquad A(W) = A(F) + \epsilon\Delta_{F,G} + \frac{\epsilon^2}{2} \Delta_{F,G}^{(2)} + \frac{\epsilon^3}{6} \Delta_{F,G}^{(3)} + \cdots .$$

For small ϵ,

$$\text{(2.10)} \qquad A(W) \doteq A(F) + \epsilon\Delta_{F,G} .$$

This exhibits the role of $\Delta_{F,G}$. It helps give an indication of the robustness of the corresponding M-estimator as G is allowed to range over possible contaminating situations. Of course, in comparing different estimators, the A(F) contribution to (2.10) will dominate when ϵ is small. For ϵ moderate, $\Delta_{F,G}$ assets itself, so that an estimator with a low asymptotic variance at the base distribution F could be "beaten" (in an overall sense) by an estimator with higher A(F) but lower $\Delta_{F,G}$. For large ϵ we would look at the whole functional

$$(2.11) \quad A(W) = A[(1 - \epsilon)F + \epsilon G] = \frac{(1 - \epsilon)\int \psi^2 dF + \epsilon \int \psi^2 dG}{[(1 - \epsilon)\int \psi' dF + \epsilon \int \psi' dG]^2} .$$

Note that the expression (2.14) follows also from taking the Taylor expansion of $A(W) = A[(1 - \epsilon)F + \epsilon G]$ viewed as a function of ϵ.

We see in (2.9) how the first term $\Delta_{F,G}$ is important in approximating the asymptotic variance A(W). The quantity $\Delta_{F,G}$ is subject to another interpretation, as follows. If we put $G = \delta_x$, a point mass at x, then Δ_{F,δ_x} is the "influence curve" (Hampel [6],[7]) of A(F). Considered as a function of x, it shows what happens to the estimator $A(F_n)$ of $A(F)$ if an additional observation is "thrown in" at the point x. This interpretation is of little relevance here, however, since our objective is to find T such that $T(F_n)$ is robust for estimation of $\theta = T(F_\theta)$ rather than such that $A(F_n)$ is robust for $A(F_\theta)$.

3. COMPARISON OF THE KNOWN ESTIMATORS H15, 25A and AMT.

Huber [8] developed a theory of robust estimation for the location parameter of a normal distribution. In the classical formulation, the observations have distribution $\Phi(x - \theta)$, and an estimator which is best in the sense of asymptotic variance is the maximum likelihood estimator.

This is the M-estimator corresponding to $\psi(x) = -\Phi''(x)/\Phi'(x) = x$, or the sample mean. In Huber's formulation, the observations have distribution $F(x - \theta)$, where F is assumed to belong to a class C of distributions "close" to Φ, and an M-estimator is sought which minimaxes $A(F)$ in C. That is, the "robust" M-estimator sought is the one which minimizes $\sup_{F \in C} A(F)$. In particular, Huber considers, for fixed choice of ϵ, $0 < \epsilon < 1$, the class

(3.1) $$C = \{F \mid F = (1 - \epsilon)\Phi + \epsilon G,\ G\ \text{symmetric}\}.$$

The symmetry restriction makes the location parameter θ identifiable. For this class, the minimax solution is given by the M-estimator corresponding to

(3.2) $$\psi(x) = -\psi(-x) = \begin{cases} x\ , & 0 \le x \le k \\ k\ , & x > k, \end{cases}$$

where k is determined by ϵ through

$$\frac{\epsilon}{1 - \epsilon} = \frac{2}{k}\Phi'(k) - 2\Phi(-k).$$

This estimator may be characterized as the maximum likelihood estimator for θ in the location model based on the distribution $F^* \epsilon C$ with smallest Fisher information $I(F) = \int (F''/F')^2 dF$. In terms of ψ functions, this robust (minimax) M-estimator modifies the classical $\psi(x) = x$ by truncation for $|x|$ sufficiently large. The solution $\hat{\theta}$ is a form of Winsorized mean.

Hampel [6], [7] suggested a modification designed to satisfy certain qualitative criteria - low *gross-error-sensitivity*, high *breakdown point*, small *local-shift-sensitivity*, and a not too distant *rejection point*. Basically, he added to the ψ function (3.2) a descending line segment, producing

(3.3) $$\psi(x) = -\psi(-x) = \begin{cases} x\ , & 0 \le x < a, \\ a\ , & a \le x < b, \\ a\left(\dfrac{c - x}{c - b}\right), & b \le x < c, \\ 0\ , & x > c. \end{cases}$$

This estimator has the advantage of completely rejecting outliers while giving up very little efficiency (compared to the "Hubers") at the normal.

In the Princeton Monte Carlo study (Andrews *et al.* [1]) is found a smoothed version of the "Hampel" (3.3), given by

$$(3.4) \qquad \psi(x) = - \psi(-x) = \begin{cases} \sin ax & , \ 0 \le x < \frac{\pi}{a} \\ 0 & , \ x > \frac{\pi}{a} \end{cases} .$$

Several estimators of the Huber type (3.2), and one of the Hampel type (3.3), were entered in the Princeton study and found to be quite "robust" in a broad sense. In general, the "descenders" tended to outperform the Hubers whenever the contaminating distributions were long-tailed and the amount of contamination (the value of ϵ) at least moderate. We shall compare three specific estimators: H15, the Huber with $k = 1.5$; 25A, the Hampel with $a = 1.69$, $b = 3.04$, $c = 6.4$; and AMT, the sine function with $a = .7062$. In Figure A we have plotted the "influence curves" of these three estimators. For a functional T defined by (1.1), and differentiable according to Definition 2.1, the influence curve is the derivative evaluated at $G = \delta_x$:

$$(3.5) \qquad IC_{T,F}(x) = \frac{d}{dt} T[F + t(\delta_x - F)] \big|_{t = 0} = \frac{\psi(x)}{\int \psi' dF} ,$$

so that the influence curve of an M-estimator is conveniently just the ψ function times a coefficient of proportionality. (The general interpretation of an influence curve was mentioned at the end of Section 2.)

(In the Princeton study, scale and location parameters were estimated simultaneously, whereas here we are considering for simplicity the location problem with scale assumed known and set equal to 1. The use of the median deviation as a scale estimate made 25A and AMT have slightly different forms in the Princeton study.)

We now center attention on the neighborhood of Φ defined by (3.1). One choice for $F = (1 - \epsilon)\Phi + \epsilon G$ is given by $G = \delta_{|x|}$, where $\delta_{|x|}$ places probability $\frac{1}{2}$ each at x and -x. This distribution is somewhat extreme but helps us understand what happens to the asymptotic variance if contamination consists of an observation thrown in randomly at x or -x. Using the results of Section 2, we have

(3.6) $$A(F) \doteq A(\Phi) + \epsilon\Delta_{\Phi,\delta_{|x|}} \quad ,$$

where

$$\Delta_{\Phi,\delta_{|x|}} = A(\Phi)\left(1 - \frac{2\int\psi'd\delta_{|x|}}{\int\psi'd\Phi}\right) + \frac{\int\psi^2 d\delta_{|x|}}{[\int\psi'd\Phi]^2}$$

(3.7) $$= A(\Phi)\left(1 - \frac{2\psi'(x)}{\int\psi'd\Phi}\right) + \frac{\psi^2(x)}{[\int\psi'd\Phi]^2} \quad .$$

In Figure B we plot $\Delta_{\Phi,\delta_{|x|}}$ versus x for each estimator. It is seen that H15 is superior for $x \leq 4$, but AMT and 25A come into their own for $x > 4$. For $x > 4.46$ and $x > 6.4$, the $\Delta_{\Phi,\delta_{|x|}}$ values for AMT have dropped to 1.042 and 1.026, their respective asymptotic variances at Φ, whereas the value of $\Delta_{\Phi,\delta_{|x|}}$ for H15 remains *constant* at 4.03. The minimax properties of H15, as well as the pitfalls of such a principle, are well illustrated by Figure B.

Another choice for the contaminating distribution is N_x, normal with mean 0 and variance x^2. This contamination model was used in the Princeton study for x = 3 and x = 10, and in terms of $\Delta_{\Phi,G}$ curves is fairly representative of other diffuse contaminations (for example, the Laplace distribution). Now, for $F = (1 - \epsilon)\Phi + \epsilon N_x$, we have

(3.7) $$A(F) \doteq A(\Phi) + \epsilon\Delta_{\Phi,N_x} \quad ,$$

where

$$\Delta_{\Phi,N_x} = A(\Phi)\left(1 - \frac{2\int\psi'dN_x}{\int\psi'd\Phi}\right) + \frac{\int\psi^2 dN_x}{[\int\psi'd\Phi]^2} \quad .$$

Notice that the value of Δ_{Φ,N_x} for a particular x is just the integral of $\Delta_{\Phi,\delta_{|x|}}$ with respect to the measure N_x. The plots of Δ_{Φ,N_x} in Figure C speak even more highly in favor of the descenders. It is seen that H15 is clearly outperformed for x > 2.5 and is competitive only for x < 2.5.

Although the Hubers achieve optimality in a certain (minimax) sense of robustness, it was discovered empirically in the Princeton study that at "typical" cases of contaminating distributions the Hubers are actually outperformed by certain other estimators, particularly "descenders." Figure B and C show analytically the basis of this phenomenon. For the Hubers, A(F) tends to remain at levels near $\sup_{F \in C} A(F)$ over a wide range of F∈C; the opposite is the case for 25A and AMT.

These $\Delta_{\Phi,G}$ plots not only help us compare old estimators, but also allow us to set up criteria for building new estimators. As a first criterion, it seems reasonable to require all robust M-estimators to have descending ψ functions. Secondly, the peak value of $\Delta_{F,G}$, as G varies, should be kept low while maintaining efficiency at the base distribution. Thirdly, the comparison of 25A and AMT in Figures A and B suggests that smoothing at the bends of the 25A ψ function should improve the quality of the estimator. Accordingly, thinking of $G = \delta_{|x|}$ as the "worst" kind of distribution (in the sense of producing irregularity), we formulate a *modified minimax* objective: find an estimator with a smooth and descending ψ function having relatively low asymptotic variance at the base distribution and having relatively minimal peak value of $\Delta_{F,\delta_{|x|}}$. In Section 4 we develop some new estimators which tend to fulfill this objective.
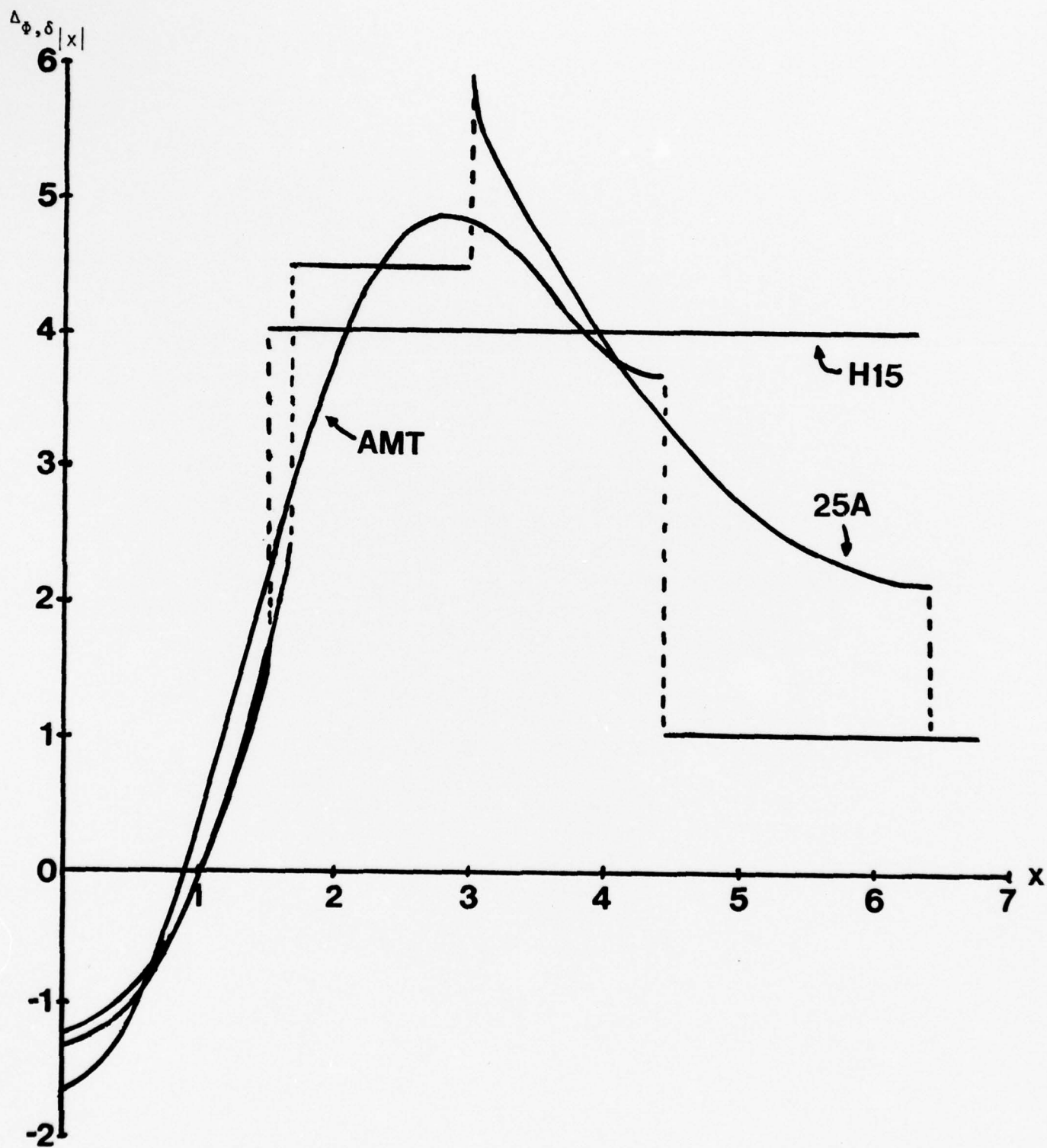
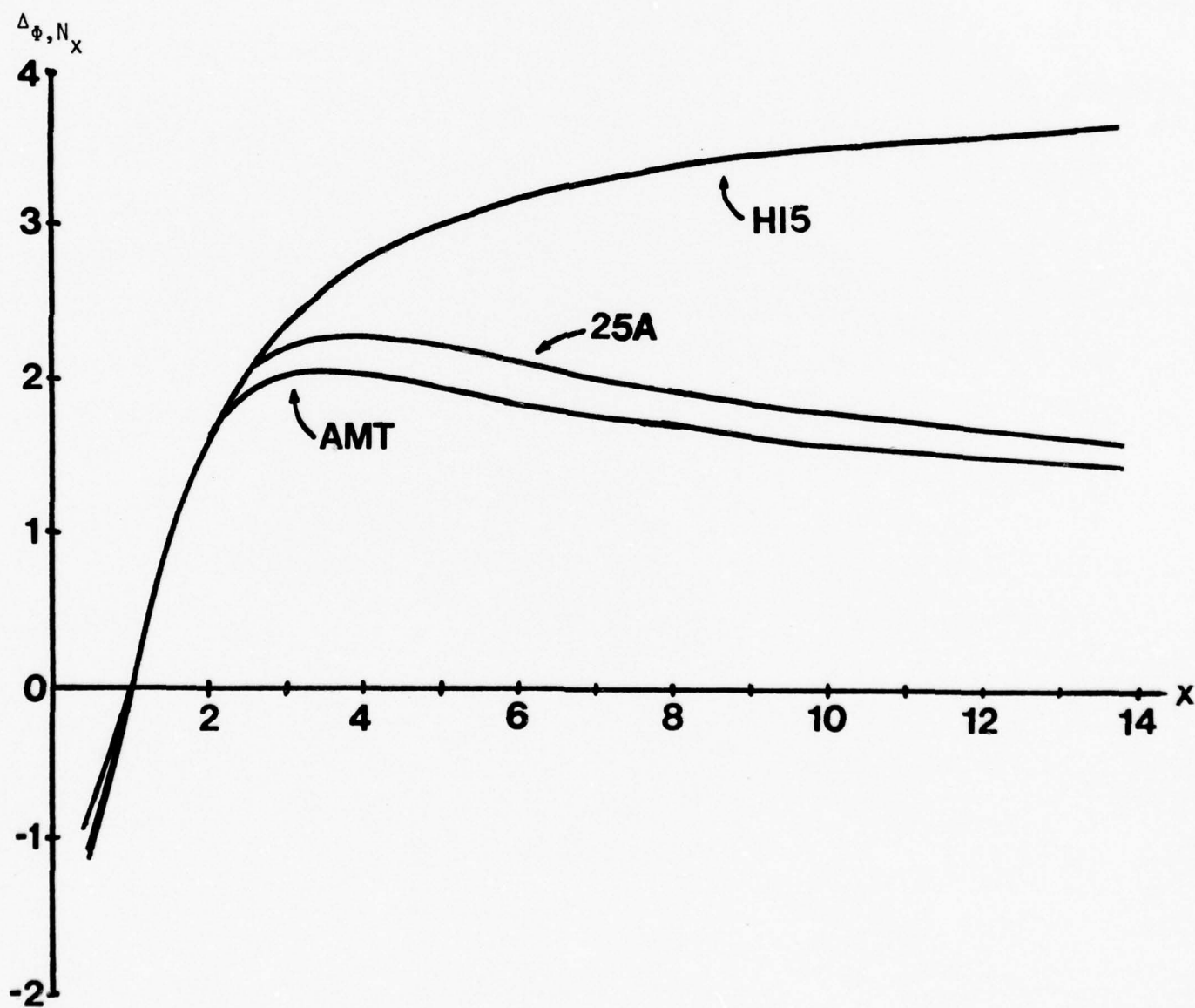A. Influence Curves for H15, 25A, and AMT at Standard Normal
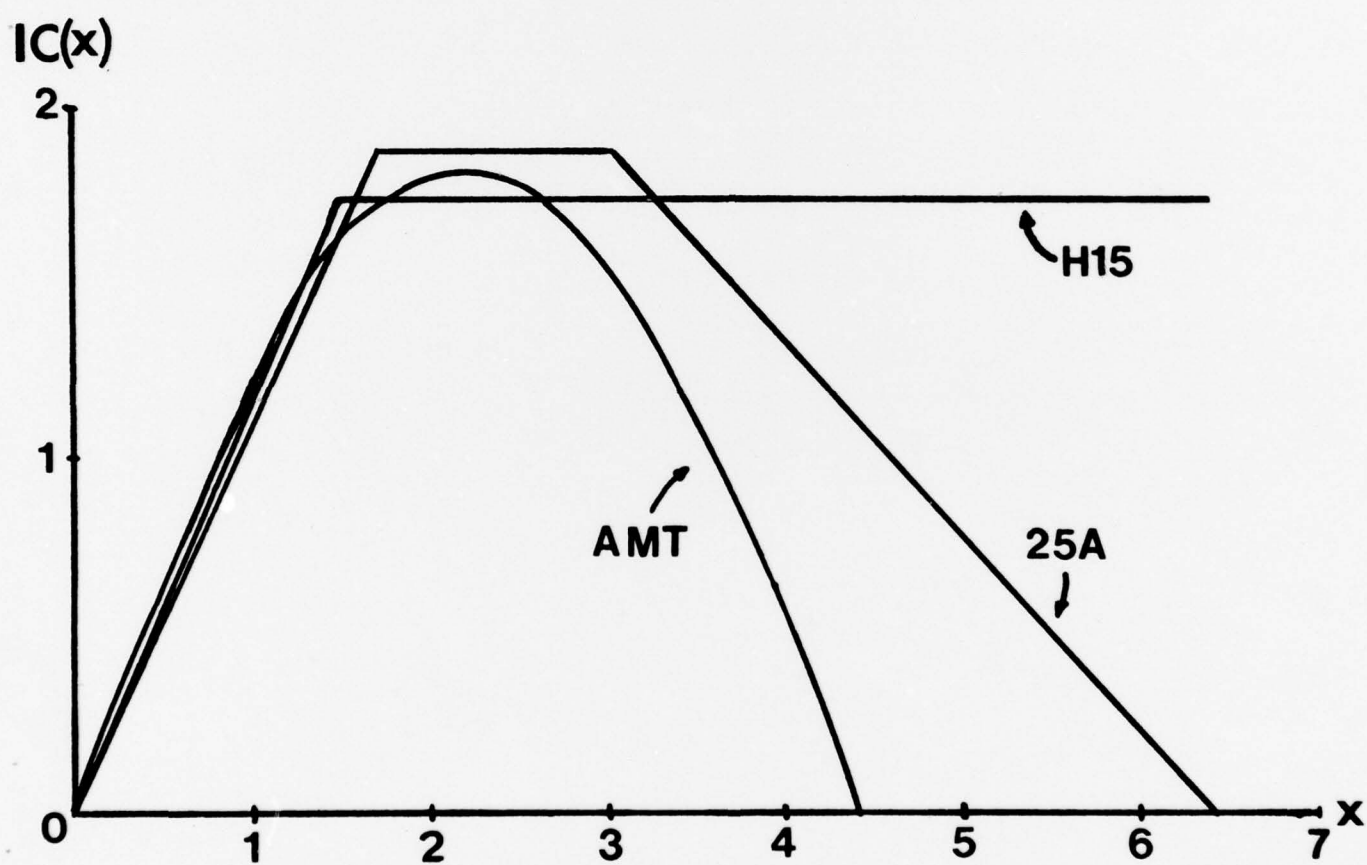
NOTE: $IC(x) = \psi(x)/\int \psi'(x)d\Phi(x)$.

B. $\Delta_{\Phi,\delta}|x|$ for H15, 25A, and AMT

C. Δ_{Φ,N_X} for H15, 25A, and AMT

4.   DEVELOPMENT OF NEW ESTIMATORS H15D, 25AR AND QC45

We have seen in Section 3 that H15 needs a descending tail and 25A a smoother descending tail. We can produce some new estimators by modifying these old estimators. We introduce H15D with ψ function

$$(4.1) \qquad \psi(x) = -\psi(-x) = \begin{cases} x & , \quad 0 \leq x \leq 1.5, \\ 1.5 & , \quad 1.5 < x \leq 2, \\ 1.5 \sin(.5236x + .5236), & 2 < x \leq 5, \\ 0 & , \quad x > 5, \end{cases}$$

and 25AR with ψ function

$$(4.2) \qquad \psi(x) = -\psi(-x) = \begin{cases} x & , \quad 0 \leq x \leq 1.69, \\ 1.69 & , \quad 1.69 < x \leq 3.04, \\ 1.69 \sin(.4675x + .1596), & 3.04 \leq x \leq 6.4, \\ 0 & , \quad x > 6.4. \end{cases}$$

The choice of a sine function for the descending part is rather arbitrary, as many other similar functions would suffice. One might think it important to smooth the first bend of the ψ function, but we have explored this and found that it does not help in minimizing $\Delta_{F,\delta_{|x|}}$ and definitely increases the asymptotic variance at ϕ.

We now exhibit a totally new estimator. In lieu of sine functions, we consider a fairly simple one parameter family of *quadratic* functions given by

$$(4.3) \qquad \psi(x) = -\psi(-x) = \begin{cases} cx - x^2 & , \quad 0 \leq x \leq c, \\ 0 & , \quad x > c. \end{cases}$$

By computing the asymptotic variance and $\Delta_{\phi,\delta_{|x|}}$ curves for several values of c, we find a range of c values yielding robust but fairly efficient M-estimators. It is interesting to note that as c increases the asymptotic variance at ϕ decreases, but the peak of the $\Delta_{\phi,\delta_{|x|}}$ curve is raised. This is not surprising since as $c \to \infty$ this estimator approaches $\overline{X}$, but it does point out the compromise

one always must make between efficiency at the normal and high levels
of protection against non-normality. In plotting $\Delta_{\Phi,\delta}|x|$ curves, one
defect in the above quadratic family was discovered. A second peak
in the $\Delta_{\Phi,\delta}|x|$ curve is caused by the ψ function of the estimator
descending to 0 too rapidly; more precisely, the contribution to (3.6)
from the quantity $\psi'(x)/\int\psi'd\Phi$ is too large. Hence, instead of (4.3),
we adopt the family of ψ functions

$$(4.4) \qquad \psi(x) = -\psi(-x) = \begin{cases} cx - x^2 , & 0 \le x \le d, \\ -e(x - f) , & d \le x \le f, \\ 0 , & x \ge f, \end{cases}$$

where d, e and f are chosen so that the influence curve of ψ, $\psi(x)/\int\psi'd\Phi$,
has slope 1 for $x \ge d$ and ψ' is continuous. In particular, we take
c = 4.5, which makes d = 3.7, e = 2.9 and f = 4.72, and we call the
corresponding M-estimator QC45. Its ψ function is thus

$$(4.5) \qquad \psi(x) = -\psi(-x) = \begin{cases} 4.5x - x^2 , & 0 \le x \le 3.7, \\ -2.9(x - 4.72), & 3.7 \le x \le 4.72, \\ 0 , & x > 4.72. \end{cases}$$

A plot of the influence curves of H15D, 25AR and QC45 is presented
in Figure D. In Section 5 we show how these three estimators compare
with those examined in Section 3.

D. Influence Curves for H15D, 25AR and QC45 at Standard Normal
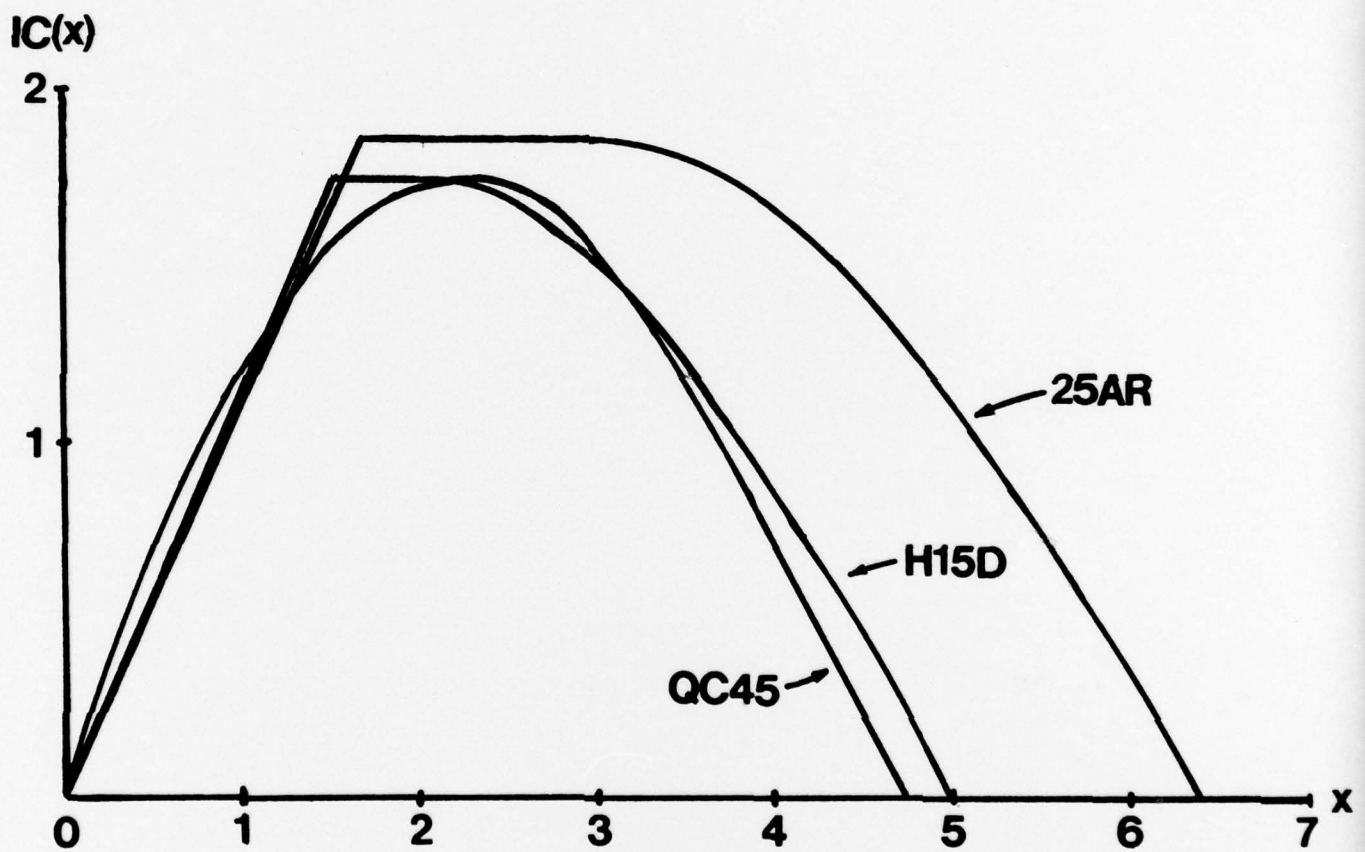
NOTE: $IC(x) = \psi(x)/\int \psi'(x)d\Phi(x)$

5.   COMPARISON OF OLD AND NEW ESTIMATORS

Three new estimators were developed in Section 4 within the guidelines set forth in Section 3. Here we investigate whether these new estimators are actually improvements over the old ones. Figures E and F give the basic comparisons among all six estimators: H15, 25A, AMT, H15D, 25AR and QC45. In Figure E the $\Delta_{\phi,\delta_{|x|}}$ curves are displayed. We see that H15D has the lowest peak value and easily beats H15 for $x \geq 3.4$. Also, 25AR has smoothed out the large spike in 25A at $x = 3.04$, but in doing so it remains higher than 25A for $6.4 \leq x \leq 3.5$. The estimator QC45 is similar to AMT in shape but is better in the range $2 \leq x \leq 4.5$. Turning to Figure F, which displays the Δ_{ϕ,N_x} curves, we can hardly distinguish between H15D, AMT and QC45, but 25A uniformly beats 25AR for $x \geq 2$, and H15, the only non-descender, is clearly outmatched by all the descenders.

Thus far we have examined only the $\Delta_{\phi,G}$ curves of these estimators. This affords an effective approximate analysis of the asymptotic variances. But what about the *exact* asymptotic variance in an ϵ-neighborhood of ϕ? Figures G, H and I plot the actual asymptotic variances for $F = (1 - \epsilon)\phi + \epsilon N_x$, based on expression (2.11), for $\epsilon = .01$, .1 and .25. Note that the scale of the ordinate is different in each figure in order to accentuate the relative standing of the estimators. For $\epsilon = .01$, the curves merely reflect the orderings of $A(\phi)$ except for the general poor behavior of H15.   In Figure H, AMT is a sure winner with H15D not far behind. Note that 25AR makes a complete switch from one of the best at $\epsilon = .01$ to the worst

(of the descenders) at ϵ = .1. The same pattern holds for ϵ = .25 in Figure I: AMT, H15D and QC45 are still close together, but all three are clearly better than 25A, and 25AR pulls further away from the others. In general, these asymptotic variance graphs reinforce the patterns already seen in Figure F. This strengthens our faith in using $\Delta_{\Phi,G}$ plots for the purpose of examining asymptotic variance. Such plots allow us to summarize in a single plot the information contained in a number of specific ϵ-neighborhood situations.

In Table 1 we give the asymptotic variances for four specific distributions: the standard normal, the Laplace (double exponential) with scale parameter 1, the t-distribution with 3 degrees of freedom, and the Cauchy distribution with scale parameter 1. Note that QC45 has the largest asymptotic variance at Φ but the smallest for the other distributions. The two best at Φ, 25A and 25AR, lose out quickly for the three longer-tailed distributions. Essentially the same pattern emerges as was seen in Figures F - I: as one moves from Φ to longer-tailed distributions, AMT, H15D and QC45 consistently outperform H15, 25A and 25AR. This result is not surprising when one notices that AMT, H15D and QC45 satisfy fairly well the modified minimax objectives proposed at the conclusion of Section 3.

Also included in Table 1 are values for some specific ϵ-contaminated normal situations. The same patterns persist. This part of the table is included primarily to coordinate with the Princeton study. In fact, one may actually compare the values for AMT, H15 and 25A with those

given in Table 5-7A, 5-7B and 5-7E of the Princeton study.   The
numbers are fairly close when one considers that our values are exact
asymptotic values, whereas the Princeton study gives Monte Carlo
results for n = 20 and also involves simultaneous scale estimation.
This illustrates how our techniques for comparison of estimators offer
a cost-saving alternative to the Monte Carlo approach.

Typically, robust M-estimators must be found iteratively. Hence,
from computational considerations, one might want to avoid sine
functions or other functions which require Taylor expansions inside
the computer. An estimator with a simple ψ function like that of
QC45 may be preferable to an estimator which barely outperforms it but
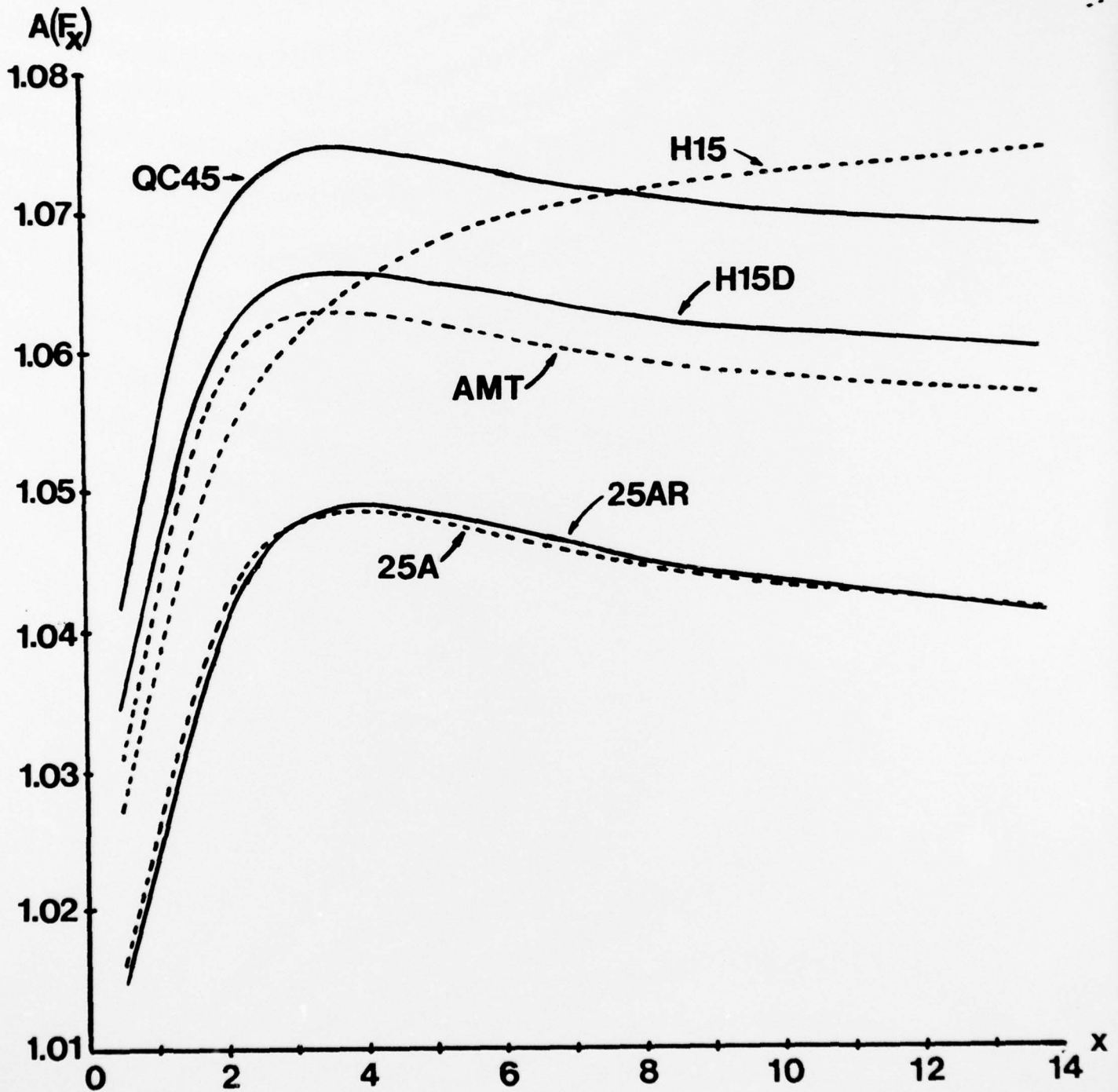is computationally worse.

E. $\Delta_{\Phi,\delta}|x|$ for H15, 25A, AMT, H15D, 25AR, and QC45
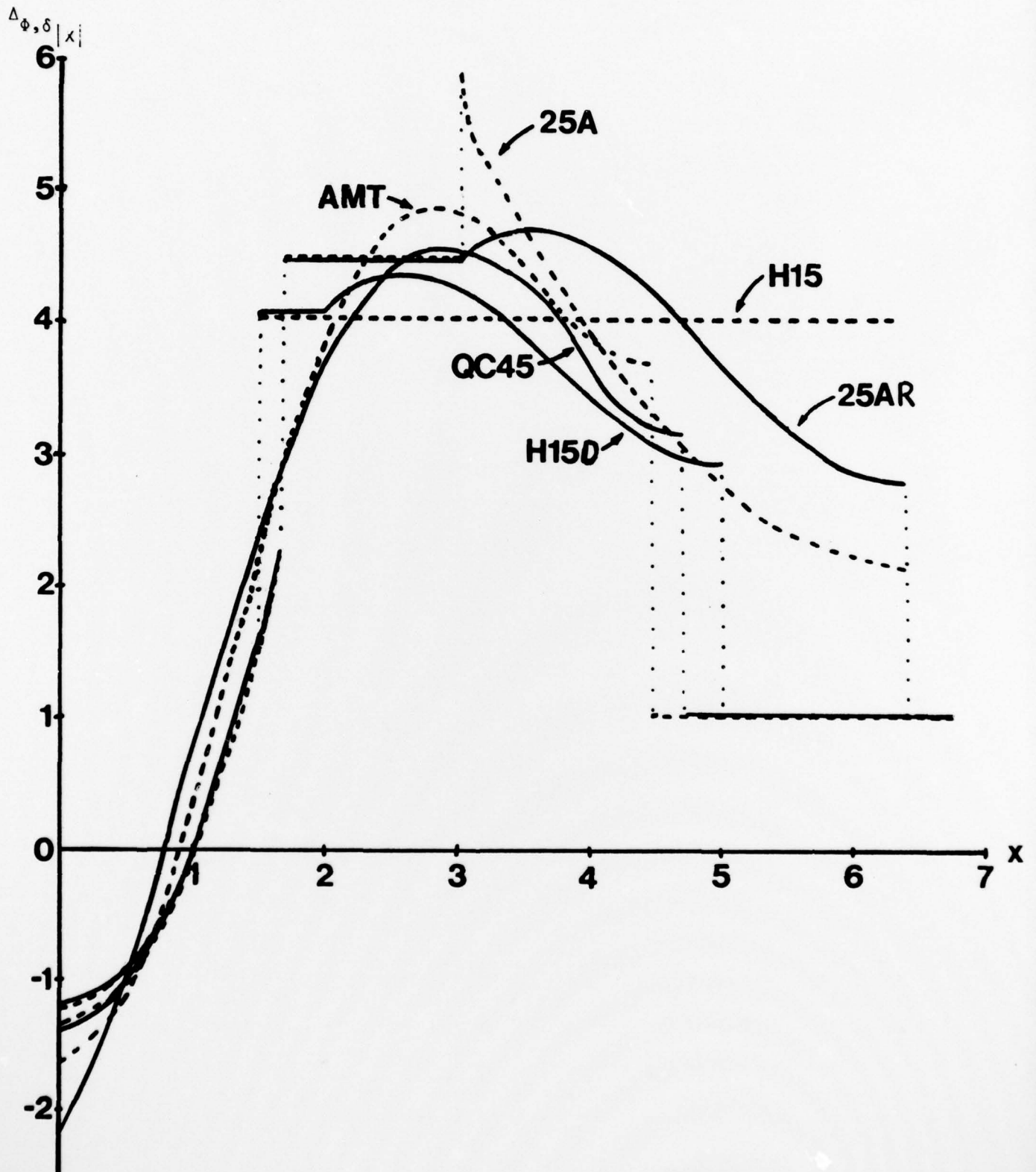
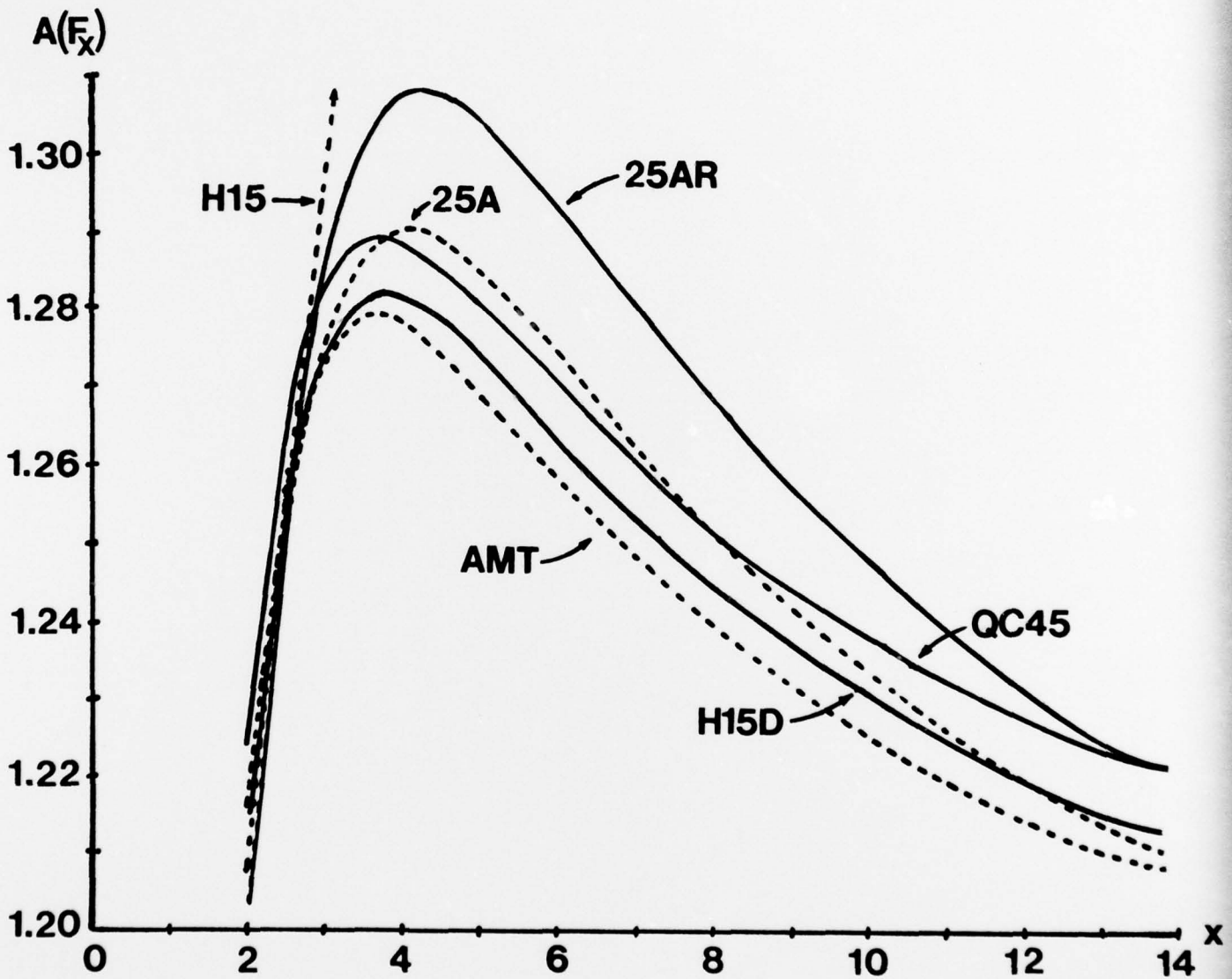F. Δ_{Φ,N_x} for H15, 25A, AMT, H15D, 25AR, and QC45
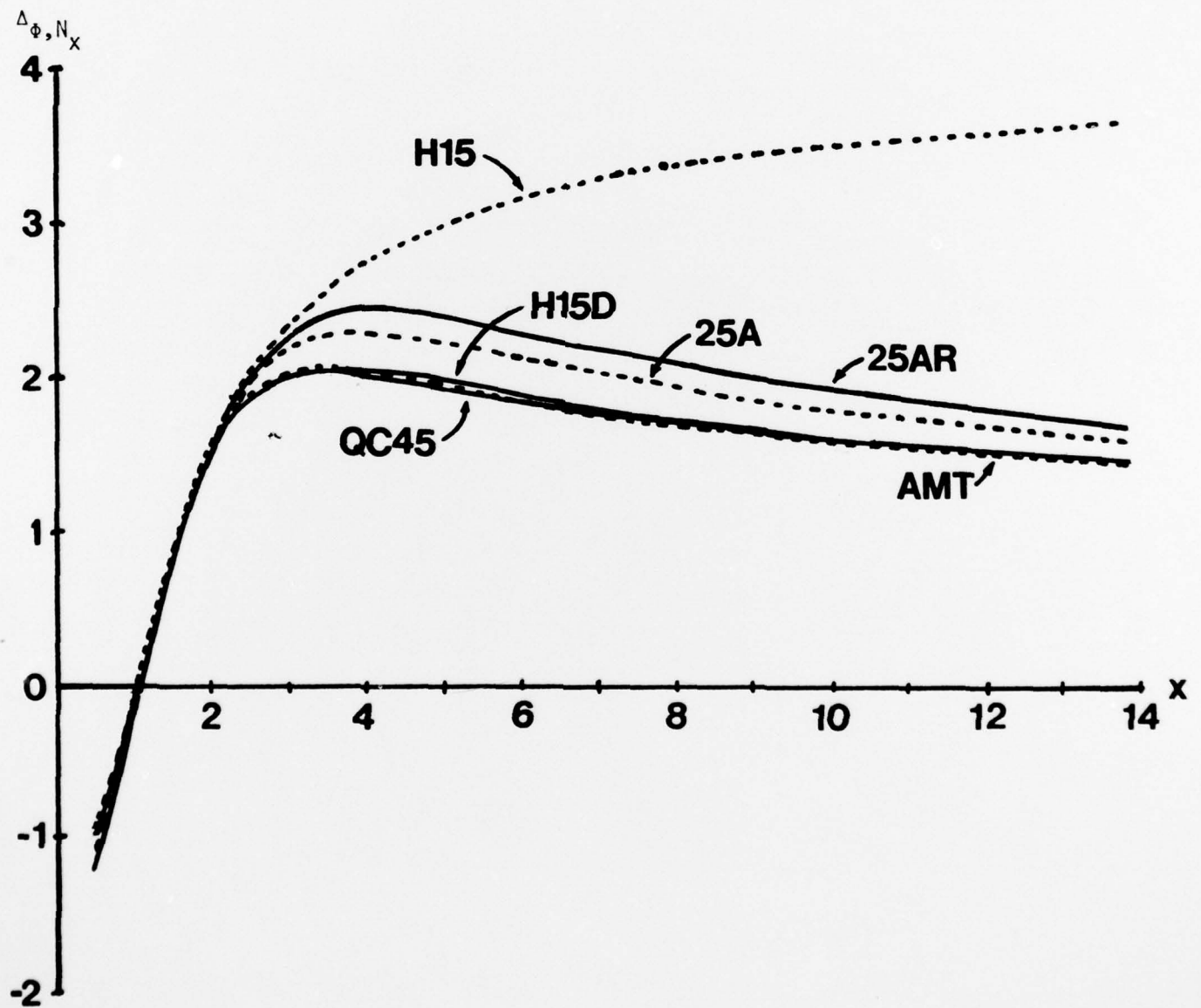
G. Asymptotic Variances of H15, 25A, AMT, H15D, 25AR, and
QC45 at $F_X = .99\Phi + .01N_X$

NOTE: $A(F_X) = [.99\int\psi^2 d\Phi + .01\int\psi^2 dN_X]/[.99\int\psi' d\Phi + .01\int\psi' dN_X]^2$.
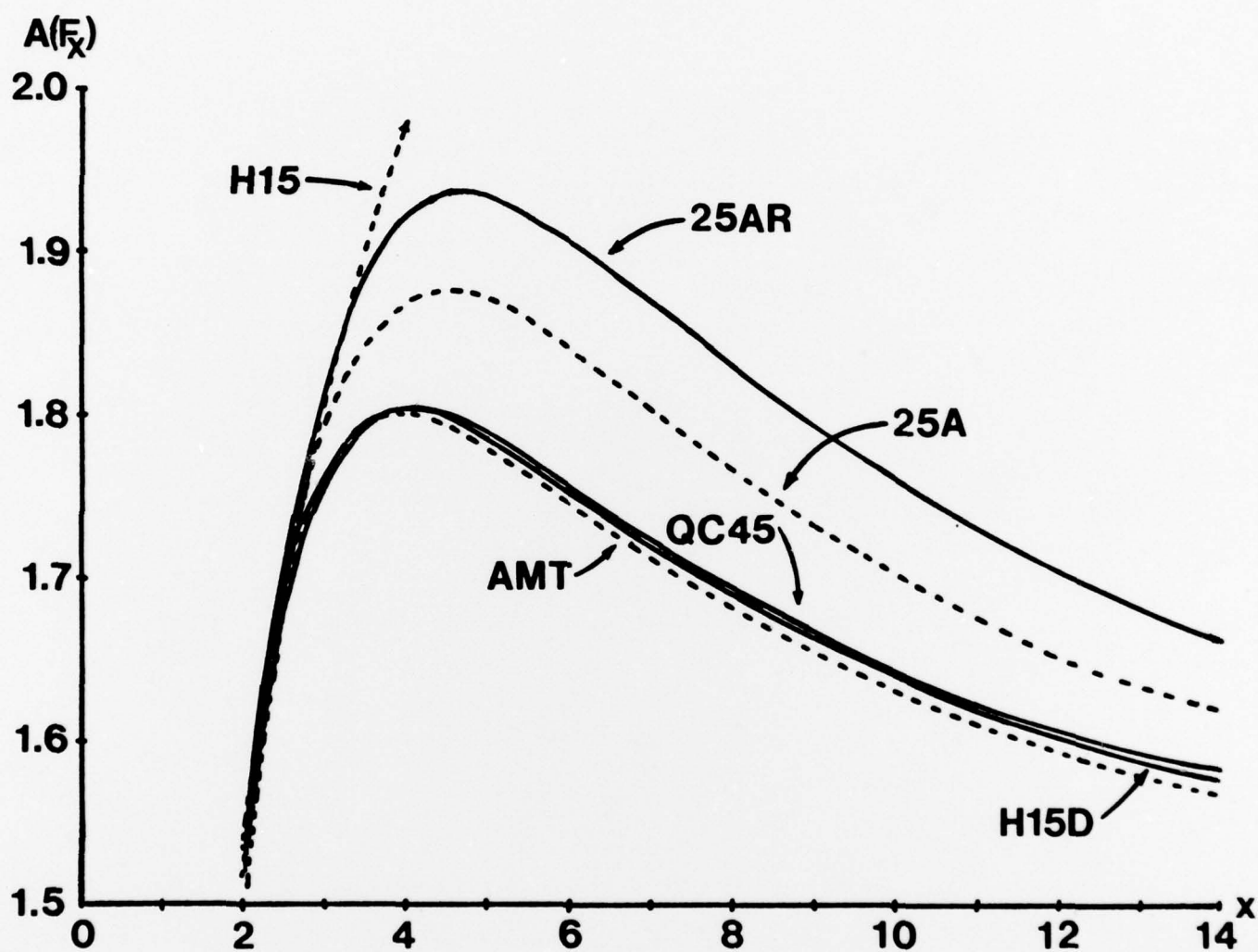
H. Asymptotic Variances of H15, 25A, AMT, H15D, 25AR, and QC45 at $F_x = .90\Phi + .10N_x$

NOTE: $A(F_x) = [.90\int\psi^2 d\Phi + .10\int\psi^2 dN_x]/[.90\int\psi' d\Phi + .10\int\psi' dN_x]^2$. For comparison, $A(F_x)$ for H15 intersects the points (4,1.347), (8,1.431) and (14, 1.469).

I. Asymptotic Variances of H15, 25A, AMT, H15D, 25AR, and
QC45 at $F_X = .75\Phi + .25N_X$

NOTE: $A(F_X) = [.75\int\psi^2 d\Phi + .25\int\psi^2 dN_X]/[.75\int\psi' dN_X + .25\int\psi' dN_X]^2$. For comparison, $A(F_X)$ for H15 intersects the points (8, 2.311) and (14, 2.474).

1. Asymptotic Variances of Selected Estimators at Selected Distributions[a]

Distribution	Estimator					
	AMT	QC45	H15	H15D	25A	25AR
Φ	1.042	1.054	1.037	1.045	1.025	1.024
Laplace	1.448	1.385	1.465	1.473	1.512	1.515
t_3	1.562	1.542	1.582	1.561	1.585	1.592
Cauchy	2.401	2.325	2.985	2.417	2.559	2.615
W(.01, 3)	1.062	1.074	1.061	1.065	1.047	1.047
W(.025, 3)	1.094	1.105	1.097	1.097	1.082	1.084
W(.05, 3)	1.150	1.160	1.160	1.152	1.142	1.147
W(.1, 3)	1.273	1.282	1.296	1.274	1.274	1.284
W(.25, 3)	1.762	1.762	1.800	1.755	1.787	1.813
W(.05, 10)	1.127	1.140	1.227	1.131	1.122	1.128
W(.1, 10)	1.226	1.238	1.448	1.231	1.234	1.249
W(.25, 10)	1.633	1.648	2.385	1.644	1.706	1.762

[a]Discussed in Sec. 5.

NOTE: The entries are the variances in the asymptotic normal distributions of the normalized estimators, when the observations have the specified distributions. The distributions are as follows: Φ is standard normal; Laplace denotes the density $f(x) = \frac{1}{2}\exp(-|x|)$; t_3 denotes the density $f(x) = 2/\sqrt{3}\pi(1 + x^2/3)^2$; Cauchy denotes the density $f(x) = 1/\pi(1 + x^2)$; $W(\epsilon, c)$ denotes the c.d.f. $(1 - \epsilon)\Phi(x) + \epsilon\Phi(x/c)$.

6. GENERALIZATIONS AND RAMIFICATIONS

We have attempted to present the functional techniques of Section 2 and their applications in the simplest situation: M-estimation in the normal location problem. Generalizations in several directions are apparent.

For example, the asymptotic variance functicnals for other types of estimators such as L-estimators (linear combinations of order statistics) or R-estimators (rank statistics) may be investigated. The techniques would be the same: find the von Mises derivative, which we denoted $\Delta_{F_o,G}$, and examine it for different choices of ϵ-neighborhoods, $F = (1 - \epsilon)F_o + \epsilon G$.

Further, it is not necessary that $F_o = \Phi$, although this is the most common base distribution. One could choose F_o to be the Laplace or some other theoretically reasonable base distribution. It is not even necessary that F_o or G be assumed symmetric as long as the asymptotic variance functional is well-defined. The technical problem of what exactly is being estimated in these asymmetric cases is discussed by Bickel and Lehmann [3]. Indeed, Collins [4] derives optimal ψ functions for such cases, and we note that one variety of solution is a "descender."

A third type of extension involves simultaneous estimation of location and scale parameters. In these cases the expression for $\Delta_{F_o,G}$ is similar to that derived in Section 2, but with an extra term due to estimation of the scale parameter. Otherwise the analysis follows as before. Note that the actual use of an M-estimator of location on real data requires simultaneous scale estimation.

Clearly the complexity of a problem is increased by the inclusion of one or more of these generalizations. Nevertheless the techniques are straight-forward and may be applied to a large class of problems.

REFERENCES

[1] Andrews, D. F. *et al.*, *Robust Estimates of Location: Survey and Advances*, Princeton: Princeton University Press, 1972.

[2] Bickel, P. J. and Lehmann, E. L., "Descriptive Statistics for Nonparametric Models. I. Introduction," *Annals of Statistics*, 3 (1975), 1038-1044.

[3] Bickel, P. J. and Lehmann, E. L., "Descriptive Statistics for Nonparametric Models. II. Location, "*Annals of Statistics*, 3 (1975), 1045-1069.

[4] Collins, John R., "Robust Estimation of a Location Parameter in the Presence of Asymmetry," *Annals of Statistics*, 4 (1976), 68-85.

[5] Filippova, A. A., "Mises' Theorem on the Asymptotic Behavior of Functionals of Empirical Distribution Functions and Its Statistical Applications," *Theory of Probability and Its Applications*, 7 (1962), 24-57.

[6] Hampel, Frank R., "Contributions to the Theory of Robust Estimation," Ph. D. dissertation, University of California, Berkeley, 1968.

[7] Hampel, Frank R., "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association*, 69 (June 1974), 383-393.

[8] Huber, P. J., "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35 (1964), 73-101.

[9] Huber, P. J., "Robust Statistics: A Review," *Annals of Mathematical Statistics*, 43 (1972), 1041-1067.

[10] Rao, C. R., *Linear Statistical Inference and Its Applications*, 2nd Edition, New York: Wiley, 1973.

[11] von Mises, R., "On the Asymptotic Distributions of Differentiable Statistical Functions," *Annals of Mathematical Statistics*, 18 (1947), 309-348.

REPORT DOCUMENTATION PAGE		
1. REPORT NUMBER ONR Technical Report No. 108	2. GOVT. ACCESSION	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and subtitle) Development and Comparison of M-estimators for Location on the Basis of the Asymptotic Variance Functional		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER FSU Statistics Report M380
7. AUTHOR(s) Dennis D. Boos R. J. Serfling		8. CONTRACT OR GRANT NUMBER(s) Office of Naval Research Contract No. N00014-76-C-0608 and N00014-75-C-0551
9. PREFORMING ORGANIZATION NAME AND ADDRESS Florida State University Department of Statistics Tallahassee, Florida 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Arlington, Virginia 22217		12. REPORT DATE July, 1976
		13. NUMBER OF PAGES 28
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report).

18. SUPPLEMENTARY NOTES

REPORT DOCUMENTATION PAGE

20. ABSTRACT

DEVELOPMENT AND COMPARISON OF M-ESTIMATORS FOR LOCATION
ON THE BASIS OF THE ASYMPTOTIC VARIANCE FUNCTIONAL

A new approach for comparison and modification of M-estimators is introduced and implemented. The problem considered is that of robust estimation of a location parameter. Specific attention is given to the ϵ-contaminated normal model. The analytical method introduced is based upon the asymptotic variance of the estimator, the asymptotic variance being considered as a functional over the space of distribution functions. The behavior of this functional is investigated with respect to special sub-families within a neighborhood of the "central" distribution. With respect to the normal location problem, three robust estimators from the 1972 Princeton Monte Carlo study are examined: the "Huber" H15, the "Hampel" 25A, and the sine function M-estimator AMT. Using the asymptotic variance functional analysis as both an analytical and intuitive tool, three modified estimators are suggested and developed. All six estimators are then compared at selected distributions heavier tailed than the normal. Besides its analytical and intuitive appeal, this functional approach offers a cost-saving alternative to Monte Carlo methods.